Credit Card Fraud Detection using Decision Tree Induction Algorithm

Jyoti R. Gaikwad, Amruta B. Deshmane, Harshada V. Somavanshi, Snehal V. Patil, Rinku A. Badgujar

Abstract— With the brisk advancement in the electronic commerce technology and improvements in the communication channels, fraud is scattering all over the world, ensuing in massive financial losses. In machine learning Fraud detection has been an interesting topic. In present day, the major causes of great financial losses is credit card fraud, which affect not only merchants but also individual clients too. Due to enormous raise in credit card transactions, credit card fraud has become more and more rampant in recent years. Clustering model, Gaussian mixture model, Bayesian networks are the presented methods to detect credit card fraud. In Proposed system, data mining technology, classification models based on ID3 decision trees and visual cryptography are applied on credit card fraud detection problem. Thus by the implementation of this approach in fraud detection systems, financial losses due to fraudulent transactions can be decreased more.

Index Terms— Data Mining, Credit card fraud, Credit Card Fraud Detection, E-Commerce Security, ID3 Decision Tree, Internet, online shopping, Visual Cryptography.

I. INTRODUCTION

Fraud can be defined as criminal trick aimed to result in financial or personal gain. Fraud prevention and fraud detection systems are two main mechanisms to avoid frauds and losses due to fraudulent activities. Fraud prevention is the upbeat mechanism with the goal of disabling the happening of fraud. Fraud detection systems come into play when the fraudsters go beyond to the fraud prevention systems and start a fraudulent transaction. The review of fraud domains and detection techniques can be found in Bolton and Hand (2002). Km. Lu, Sirwongwattana, and Huang (2004). Phu& Lee. Smith. and Gayler (2005). Sahin and Duman (2010). Day by day, the popularity of online shopping is rising. Credit card systems is one of the most famous fraud domains. Simple theft, application fraud, counterfeit cards, never received issue (NRI) and online fraud (where the card holder is not present) are many ways for Credit card frauds. Only the card's details are needed in online fraud, when the transaction is made remotely.

Revised Manuscript Received on November 2014.

Jyoti R. Gaikwad, Department of Computer Engineering, JSPM's Bhivarabai sawant Institute of Technology And Research, Wagholi, Pune, India.

Amruta B. Deshmane, Department of Computer Engineering, JSPM's Bhivarabai sawant Institute of Technology And Research, Wagholi, Pune, India.

Harshada V.Somavanshi, Department of Computer Engineering, JSPM's Bhivarabai sawant Institute of Technology And Research, Wagholi, Pune, India.

Snehal V. Patil, Department of Computer Engineering, JSPM's Bhivarabai sawant Institute of Technology And Research, Wagholi, Pune, India.

Rinku A. Badgujar, Department of Computer Engineering, JSPM's Bhivarabai sawant Institute of Technology And Research, Wagholi, Pune, India.

There is a rapid growth of committing fraudulent actions, because of the international availability of the web and ease with which users can hide their location and identity over Internet transactions.

II. PROBLEM DEFINITION

Credit card fraud is a sober and major growing problem in banking industries. With the advent of the rise of many web services provided by banks, banking frauds are also on the increase. Banking systems always have a strapping security system in order to detect and prevent fraudulent activities of any category of transactions. Totally eliminating banking fraud is almost unfeasible, but we can however minimize the frauds and prevent them from happening by machine learning techniques like Data Mining. It represents how to utilize these data and find useful information from data has become an urgent need for detection of fraud. Therefore, data mining technology has become an effective method for detection of fraud. Thus we are developing a fraud detection system for credit cards using decision tree induction algorithm for security using Data Mining Technique.

III. LITERATURE REVIEW

As per the survey there are various techniques of fraud detection. Tatusuya Minegishi and Ayahiko Niimi has Focused on classification learning which is analytical method of stream minig . They are concern with decision tree learnig and they analyse credit card transaction data as data stream and detect fraud use. Krishna Kumar Tripathi, Mahesh A. Pavaskar presents a survey of various techniques used in credit card fraud detection mechanism. Classification is one of the most familiar tasks in data mining. Decision trees, neural networks, logistic regression, nearest neighbors, and support vector machines are the main classification methods that currently exist. Decision trees are recognized as very powerful and attractive classification tools, mainly because they produce easily interpretable and well-organized results and are, in general, computationally efficient and capable of dealing with noisy data. Decision tree techniques build classification or prediction models based on the recursive partitioning of data, which begins with the entire body of data then splits the data into two or more subsets based on the values of one or more attributes, and then repeatedly splits each subset into finer subsets until the stopping criteria are met.

IV. IMPLEMENTATION

A. Decision Tree Induction algorithm

The decision tree is a structure that includes root node, leaf node & branch. Each internal node denotes a test on attribute, the outcome of test denotes each branch and the class label holds by each leaf node. The root node is the topmost node in the tree. The following decision tree is for concept to buy a computer, that denotes whether a customer at a company is likely to buy a computer or not. The test on the attribute represents each internal node. The each leaf node represents a class.

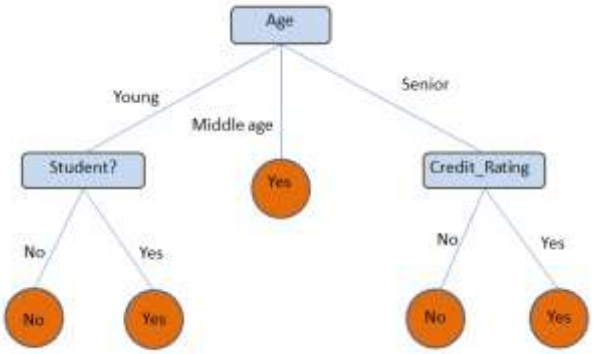

Fig. 1. A Decision Tree

B. Gray Scale Image Figures and Tables

The value of each pixel is a single sample which is known as Gray scale digital image. It carries only intensity information. The images of this sort is known as black-and-white images,are composed entirely of shades of gray, changing from black at the weakest amount to white at the strongest. To convert a color from a color space based on an RGB color model to a grayscale demonstration of its luminance, weighted sums must be calculated in a linear RGB space, that is, after the gamma compression function has been removed first via gamma extension.

For the sRGB color space, gamma extension is defined as

$$C_{\text{linear}} = \begin{cases} \frac{C_{\text{srgb}}}{12.92}, & C_{\text{srgb}} \leq 0.04045 \\ \left(\frac{C_{\text{srgb}}+0.055}{1.055}\right)^{2.4}, & C_{\text{srgb}} > 0.04045 \end{cases}$$

where Csrgb represents any of the three gamma-compressed sRGB primaries (Rsrgb, Gsrgb, and Bsrgb, each in range [0,1]) and C linear is the equivalent linear-intensity value (R, G, and B, also in range [0,1]). Then,by the three linear-intensity values luminance is calculated as a weighted sum. The sRGB color space is defined in terms of the CIE 1931 linear luminance Y, which is given by

$$Y = 0.2126R + 0.7152G + 0.0722B$$

The human vision is most sensitive to green and least sensitive to blue. To encode gray scale intensity in linear RGB, each of the three primaries can be set to equal the calculated linear luminance Y (replacing R,G,B by Y,Y,Y to get this linear grayscale).To get back to a conventional non-linear representation linear luminance typically needs to be gamma compressed. For sRGB, each of its three primaries is then fixed to the same gamma-compressed Ysrgb given by the converse of the gamma expansion above as

$$Y_{\text{srgb}} = \begin{cases} 12.92\,Y, & Y \leq 0.0031308 \\ 1.055\,Y^{1/2.4} - 0.055, & Y > 0.0031308. \end{cases}$$

In practice, because the three sRGB components are then equivalent, it is only necessary to store these values once in sRGB-compatible image formats that support a single-channel illustration.The web browsers and other softwares that identifies sRGB images will naturally creates the same rendering for such a gray scale image as it would for an sRGB image having the equal values in all three color channels.

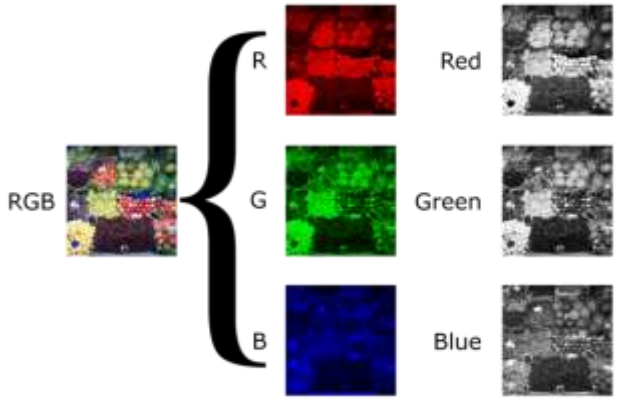

Fig. 2. A Gray Scale Image

C. Thesholding

Thresholding is the simplest method of image segmentation. For creating binary images thresholding can be used,from a gray scale image.

```
G.S= (R+G+B)/3
 R=G=B=G.S
if(G.S<th)          th-threshold value
    {
       pix=0;           indicates pure black
    }
     else
    {
       pix=0xffffff;    indicates white
    }
```

Fig. 3. Original Image

Fig. 4. Threshold Effect

D. Visual Cryptography

The visual information (pictures, text, etc.) allows through a cryptographic technique which is known as Visual cryptography to be encrypted in such a way that decryption becomes a mechanical operation that does not require a computer.

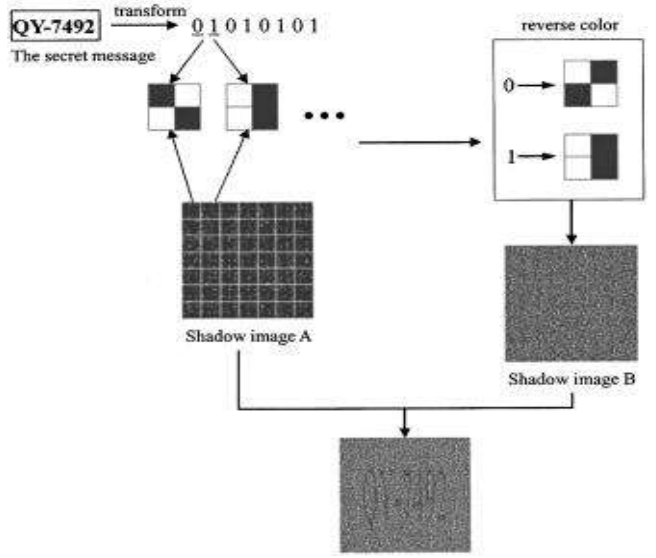

Fig. 5.Visual Cryptography

For decrypting the encrypted images visual cryptography uses the characteristics of human vision ,which is an emerging cryptography technology. It needs neither complex computation nor cryptography knowledge. It also ensures that hackers cannot perceive any clues about a secret image from individual cover images for security concerns.

E. Block Diagram

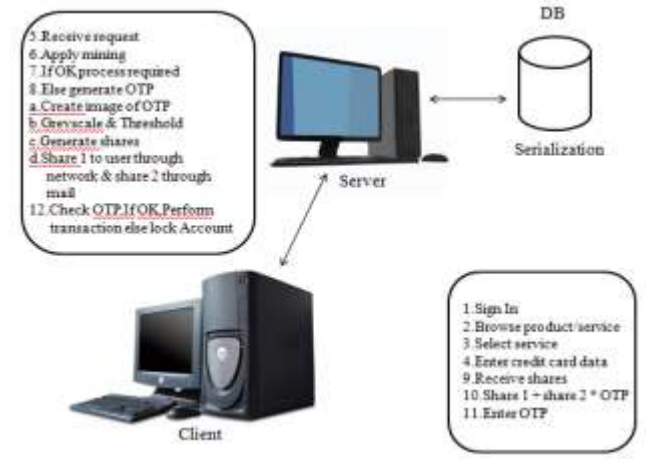

Fig. 5. Architecture of System

V. FUTURE WORK

In recent years, credit card fraud has become more and more wild. To improve merchants' risk management level in an automatic and effective way, building an accurate and easy handling credit card risk monitoring system is one of the key tasks for the merchant banks. One aim of this study is to identify the user model that best identifies fraud cases. The models are compared in terms of their performances. To improve the fraud detection system, the combination of the three presented methods could be beneficial. It is possible to use. In the future, these models can extend to use in health insurance fraud detection

VI. CONCLUSION

We propose credit card fraud detection problem for the purpose of reducing the bank's risk. With the historical profile patterns, utilize credit card fraud detection models to compare the transaction information to predict the probability of being fraudulent for a new transaction. It provides a scientific basis for the authorization mechanisms.

VII. ACKNOWLEDGMENT

We acknowledge the effort and hard work by the experts who have contributed towards the development of Credit Card Fraud Detection System. We also acknowledge the reviewers of the journal for the suggestions and modifications to improve the quality of the paper.

VIII. CONCLUSION

A conclusion section is not required. Although a conclusion may review the main points of the paper, do not replicate the abstract as the conclusion. A conclusion might elaborate on the importance of the work or suggest applications and extensions.

REFERENCES

1. Aleskerov, F., Freisteben, B.. & Rao. B. (1997). CARDWATCH:"*A neural network based data mining system for credit card fraud detection. Computational Intelligence for Financial Engineering".* 220-226.
2. Bolton. R. J. & Hand. D. J. (2002*). "Statistical fraud detection: A review. Statistical Science".* 28(3). 235-255.
3. Bradford.). P. Kunz. C.. Kohavi. R.. Brunk.C. & Bradley. C. E. (1998). "*Pruning decision tires with misclassification costs"*. In Proceedings of 10th European conference on machine learning (pp. 131-136). Berlin.
4. Brause, R, Langsdoff, T.. & Heap, M.. (1999*)." Neural data mining for credit card fraud detection"*. In Proceedings of the 1 1 th IEEE international conference on tools with artificial intelligence.
5. Breiman. L. Friedman, Isbell. R. & Stone, C. (1984). "*Classification and regression trees*". Wadsworth International Group.
6. Bradley. C_ E. (1995*). "Automatic selection of split criterion during tree growing based on node location"*. In Proceedings of 12th internationa/ conference on. machine learning (ICML-95) (pp. 73-80).
7. *Chavda, N.. Bowyer. L. & Kegelmeyer. W. (2002). SMOLT: "Synthetic minority over-sampling technique". Journal of Artificial Intelligence Research, 16, 321-357.*
8. Chen. R.. Chiu, M.. Huang, Y., & Chen. L (2004). "*Detecting credit card fraud by using questionnaire-responded transaction model based on SVMs*". In Proceedings of 1DEAL2004 (pp. 800-806) Exeter. UK.
9. Draper. R. Brod ley. C. L & Utgoff. P. (1994). "*Coal-directed classification using linear machine decision trees"*. IEEE Transactions on Pattern Analysis and Machine Intelligence. 16.888-893.

10. Drummond C.. & Hoke. R. C. (2000). "*Explicitly representing expected cost: An alternative to roc representation*". In Proc. ACM S1GKDD, int7 cool knowledge discovery and data mining (pp. 198-207).
11. Drummond. C.. & Hotta R C. (2000) "*Exploiting the cost (in)sensitivity of decision tree splitting criteria*". In Proceedings of the 17th international conference on machine reaming (pp. 239-246).
12. Duman. E..& Ozcelik, ht. H. (2011) "*Detecting credit card fraud by genetic algorithm and scatter search. Expert Systems with Applications*". 38.13057-13063.
13. Alejandro Correa Bahnsen, Aleksandar Stojanovic, Djamila Aouada and Bj¨orn Ottersten" Cost Sensitive Credit Card Fraud Detection using Bayes Minimum Risk".
14. Tatsuya Minegishi, Ayahiko Niimi "*Detection of Fraud Use of Credit Card by Extended VFDT*".
15. V.Dheepa,Dr. R.Dhanapal "*Analysis of Credit Card Fraud Detection Methods*". International Journal of Recent Trends in Engineering, Vol 2, No. 3, November 2009.
16. Krishna Kumar Tripathi, Mahesh A. Pavaskar " Survey on Credit Card Fraud Detection Methods". International Journal of Emerging Technology and Advanced Engineering,Volume 2, Issue 11, November 2012.
17. Y. Sahin and E. Duman" Detecting Credit Card Fraud by Decision Trees and Support Vector Machines".
18. Alka Herenj,Susmita Mishra "*Secure Mechanism for Credit Card TransactionFraud Detection System*". International Journal of Advanced Research in Computer and Communication EngineeringVol. 2, Issue 2, February 2013.

AUTHOR PROFILE

Jyoti R. Gaikwad, Department of Computer Engg., JSPM's Bhivarabai sawant Institute Of Technology And Research,Student of Final Year B.E.Computer.Born born at Nasikroad,(Dist.-Nasik Maharashtra) 1988.

Amruta B. Deshmane, Department of Computer Engg., JSPM's Bhivarabai sawant Institute Of Technology And Research,Student of Final Year B.E.Computer.

Harshada V.Somavanshi, Department of Computer Engg., JSPM's Bhivarabai sawant Institute Of Technology And Research,Student of Final Year B.E.Computer.

Snehal V. Patil, Department of Computer Engg., JSPM's Bhivarabai sawant Institute Of Technology And Research,Student of Final Year B.E.Computer.

Rinku A. Badgujar, Department of Computer Engg., JSPM's Bhivarabai sawant Institute Of Technology And Research, Wagholi, Pune, India